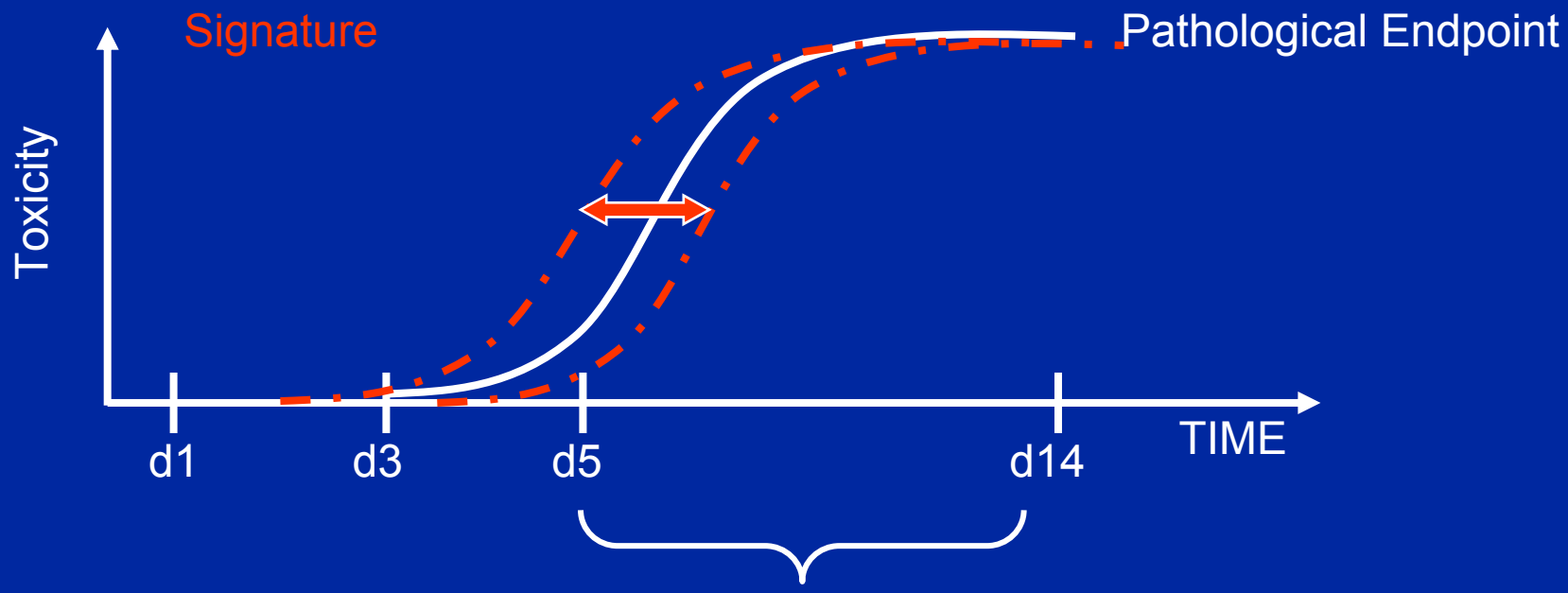
Characteristics of In Vivo and In Vitro Toxicogenomic Signatures Predictive of Toxicological Outcomes

Mark Fielden, Ph.D., DABT
Roche Palo Alto LCC

What is a Signature?

- A method of identifying the current (diagnostic) or future (predictive) phenotype induced by a compound based on multivariate data (i.e. genes)
    - In vivo: compound-dose properties (dose matters)
    - In vitro: compound properties (hazard identification)
- Typically a multivariate classification model, but may also be a profile

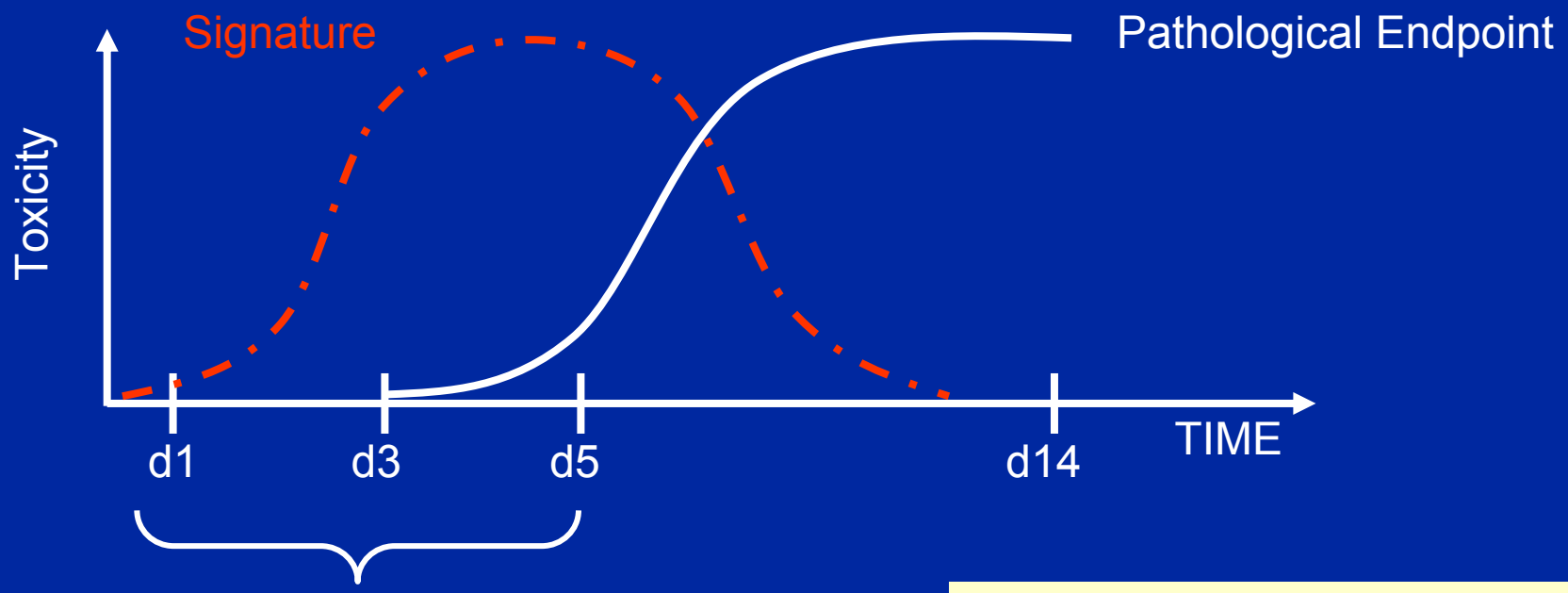
Diagnostic Signatures – Correlates with Injury, but May Have Some Predictive Utility



Biomarker coincides with injury - may be more sensitive than apical endpoint and thus predictive

Typically trained on late timepoint data when injury is observed

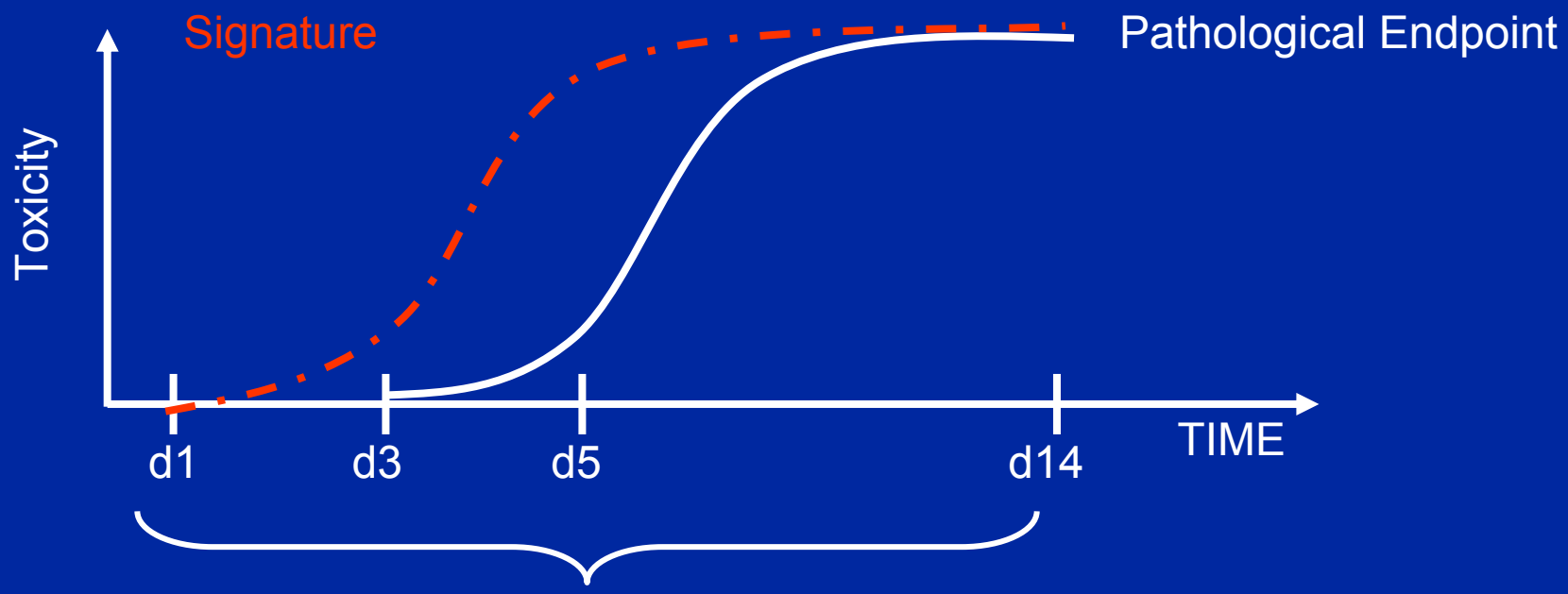
Predictive Signatures – Precedes Injury



Typically trained on early timepoint data before injury is observed

Biomarker precedes injury and may decrease upon presence of injury

Predictive Signatures – Precedes and Correlates with Injury



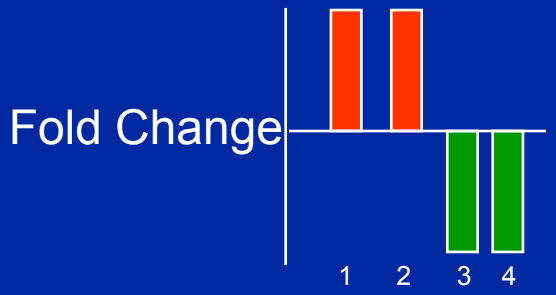
Biomarker precedes injury and is sustained during injury, resolves as injury

Signatures: Classifiers and Profiles

Linear Classification Model: $Y = cX_1 + cX_2 + cX_3 + cX_4 + B$

Non-Linear Classification Model: $Y = cX_1^{1/2} + X_2^3 + (X_3 \bullet X_4) + B$

Profile (or Pattern):

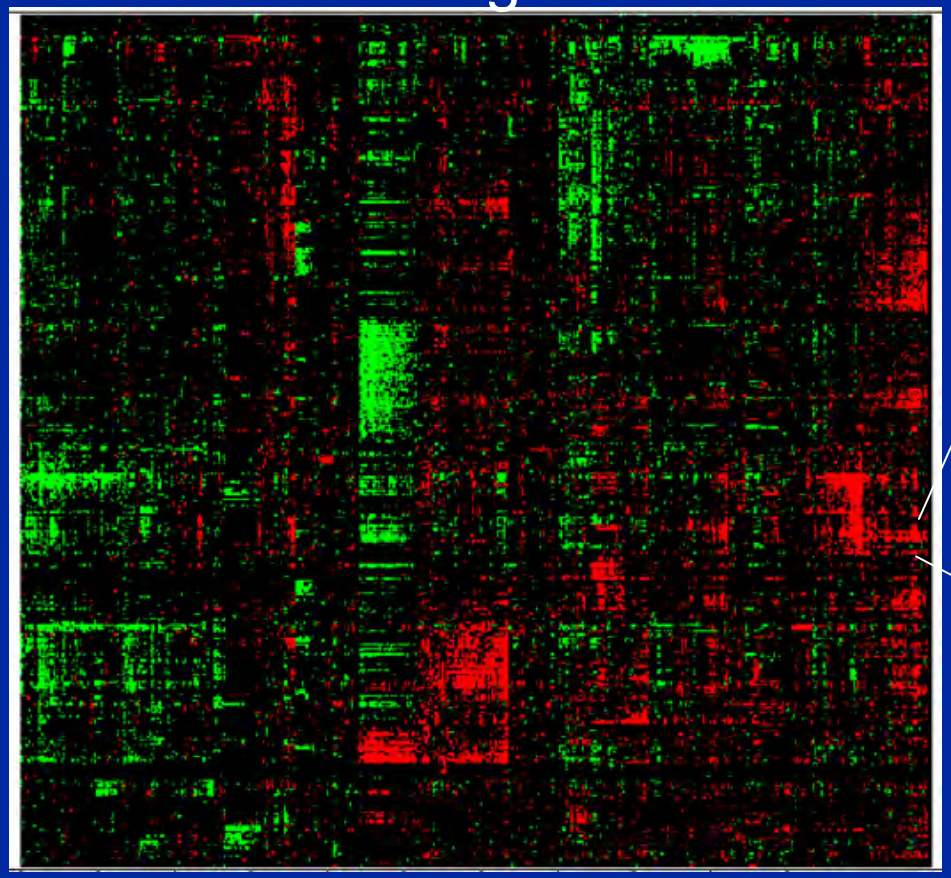


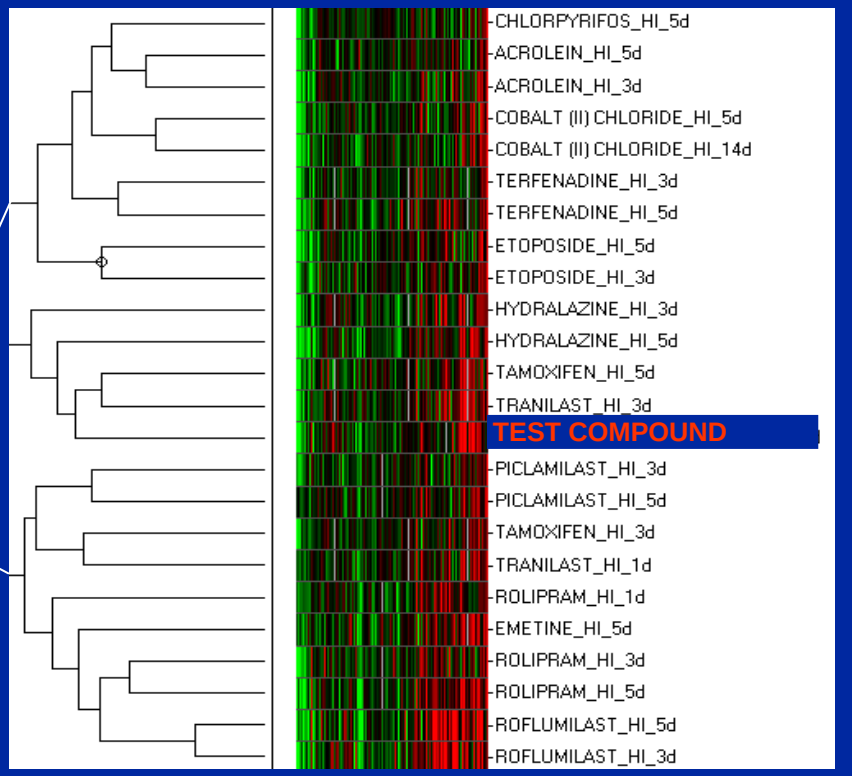
Biomarker:

Why Signatures?

1000's genes

1700 liver experiments

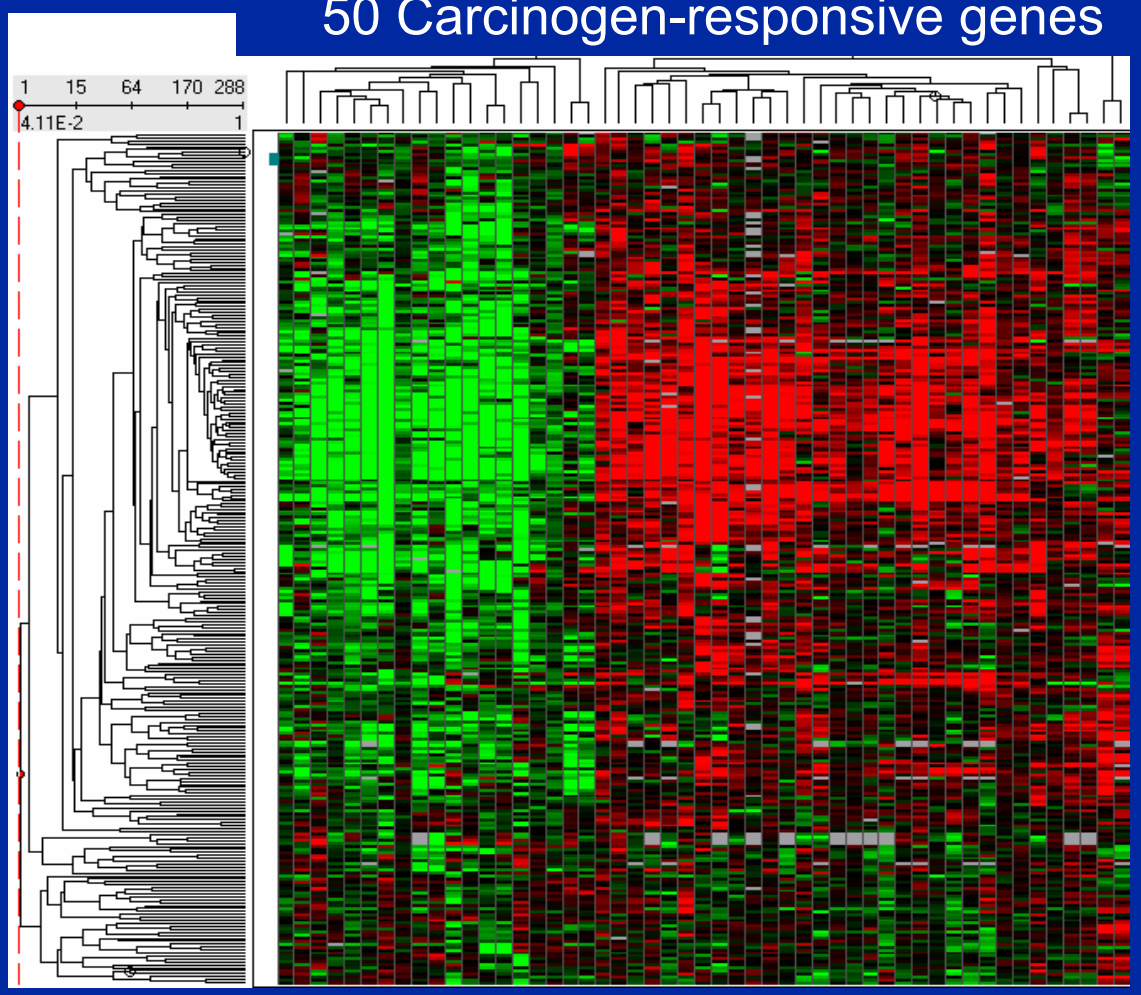




Unsupervised methods are not designed for class identification, but rather class discovery

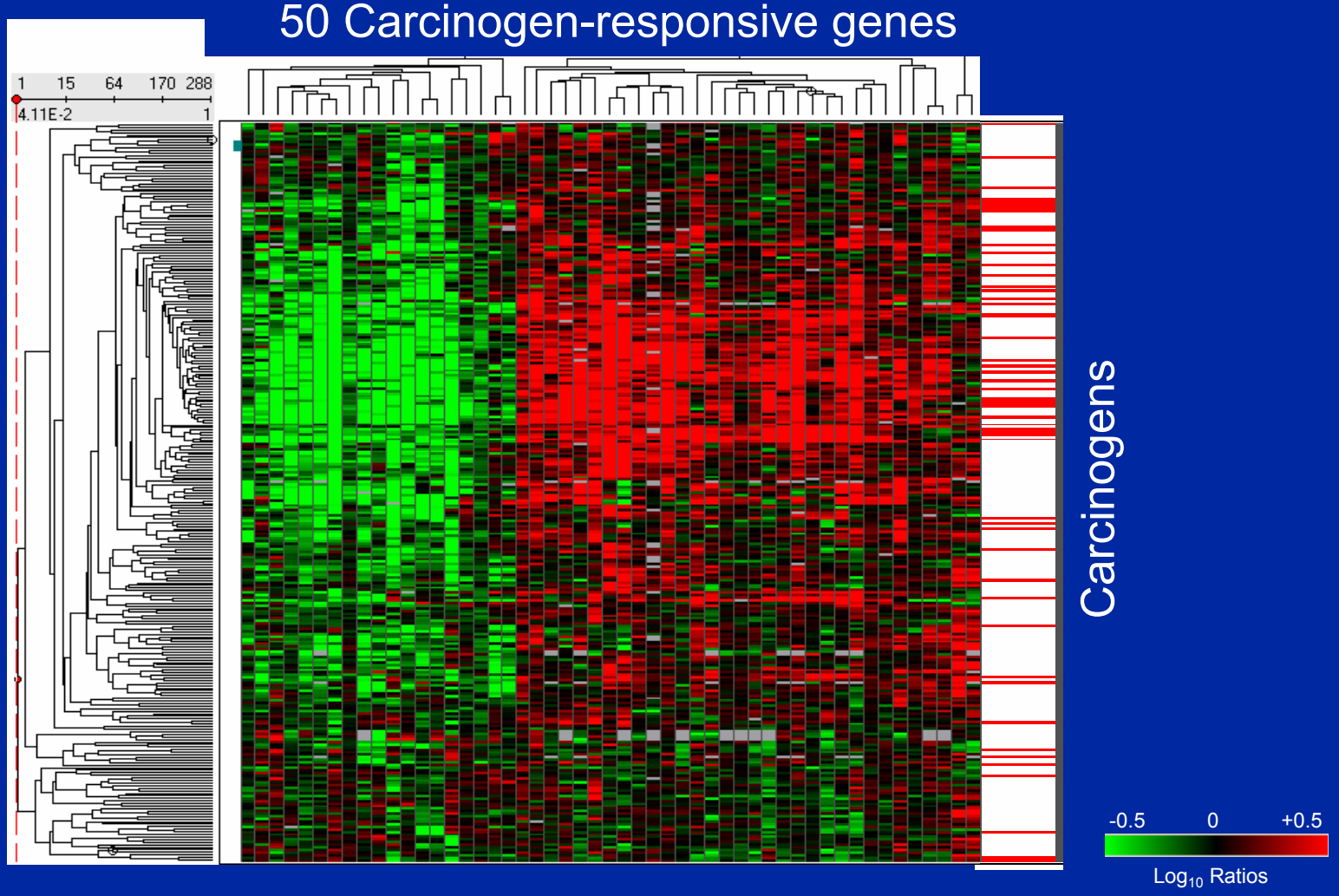
Data Provided by Iconix Biosciences

Unsupervised Methods do not Classify Complex Phenotypes, like Pathology, Very Well



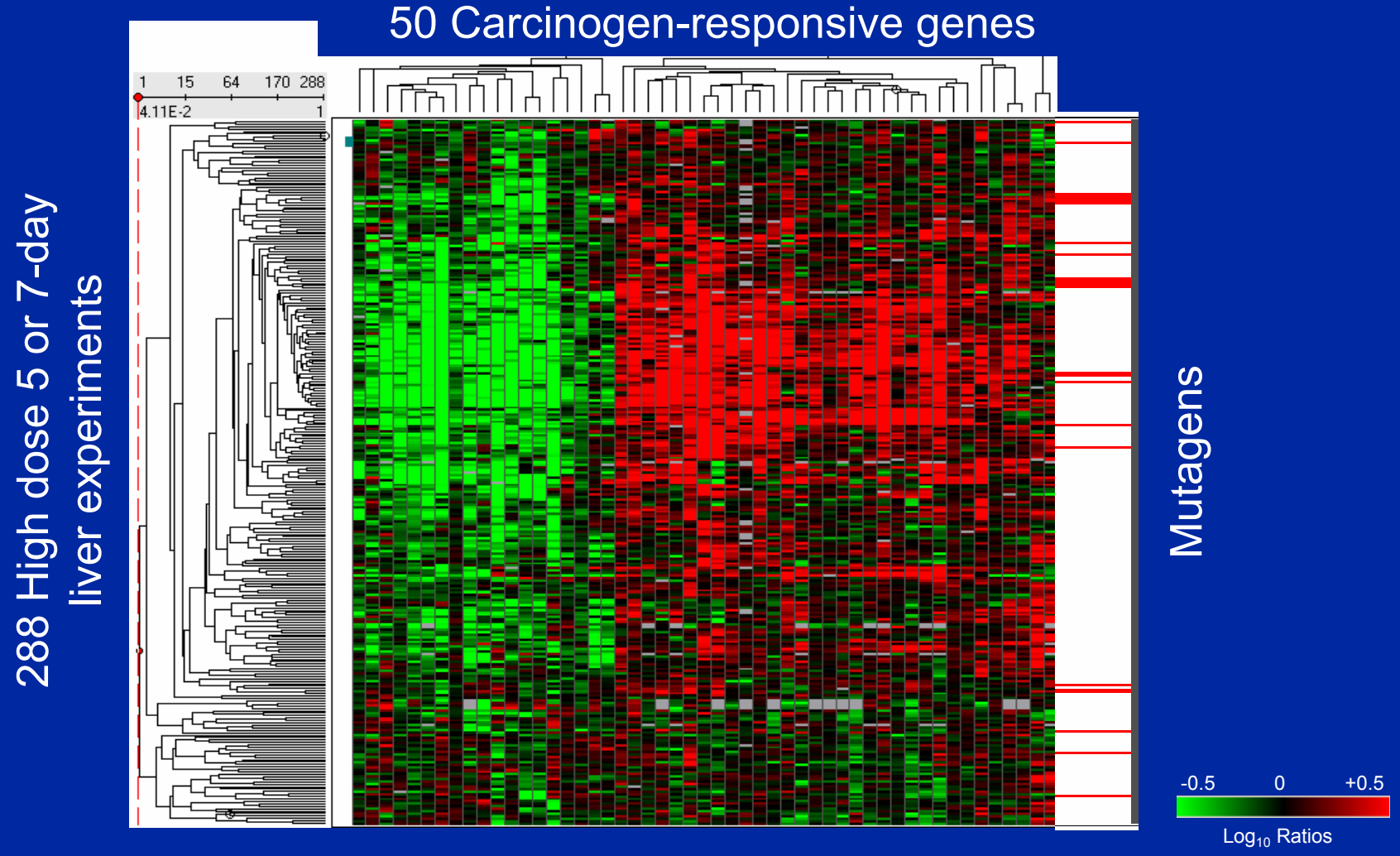
Data Provided by Iconix Biosciences

Unsupervised Methods do not Classify Complex Phenotypes, like Pathology, Very Well



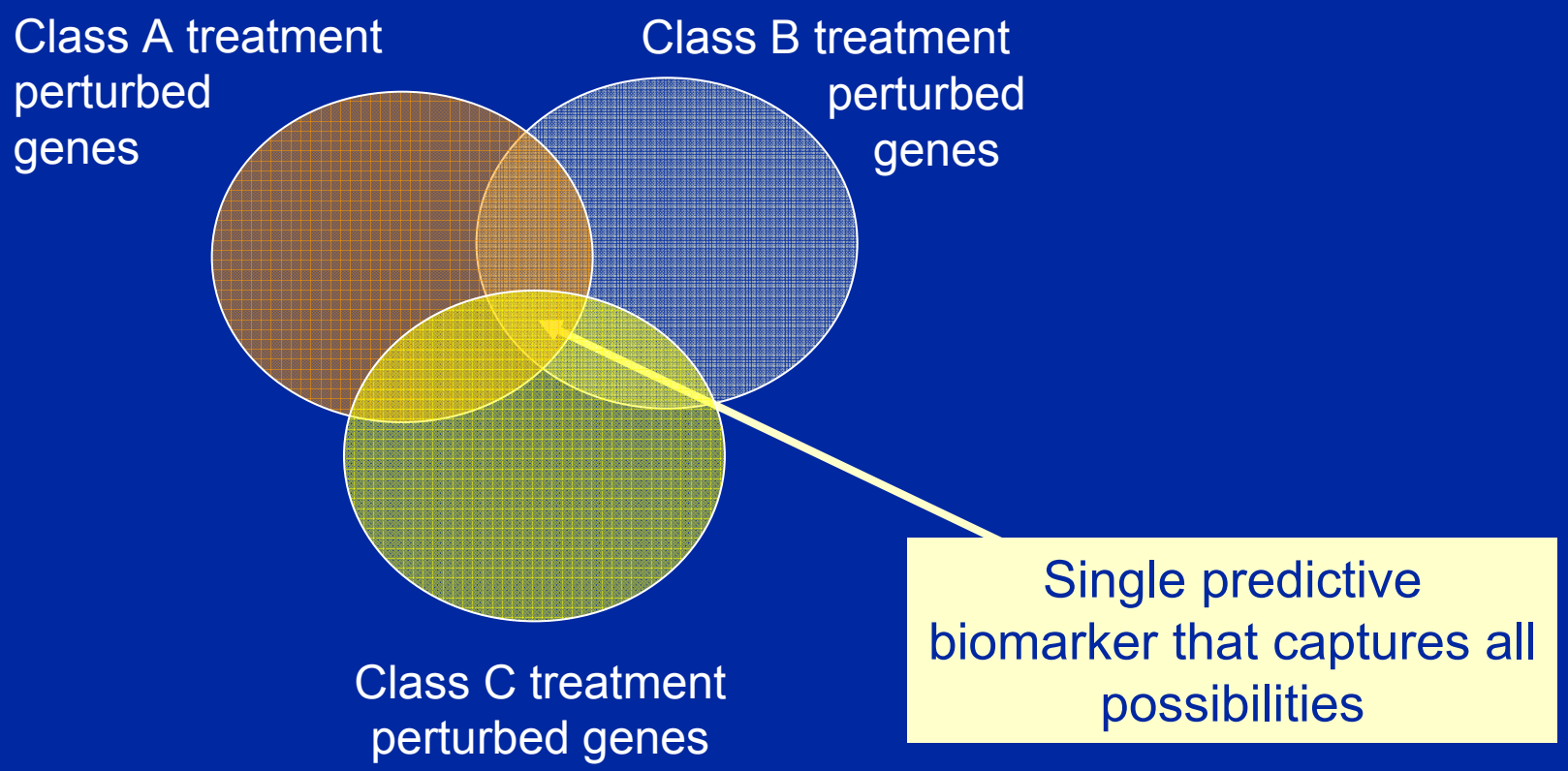
Data Provided by Iconix Biosciences

Unsupervised Methods do not Classify Complex Phenotypes, like Pathology, Very Well

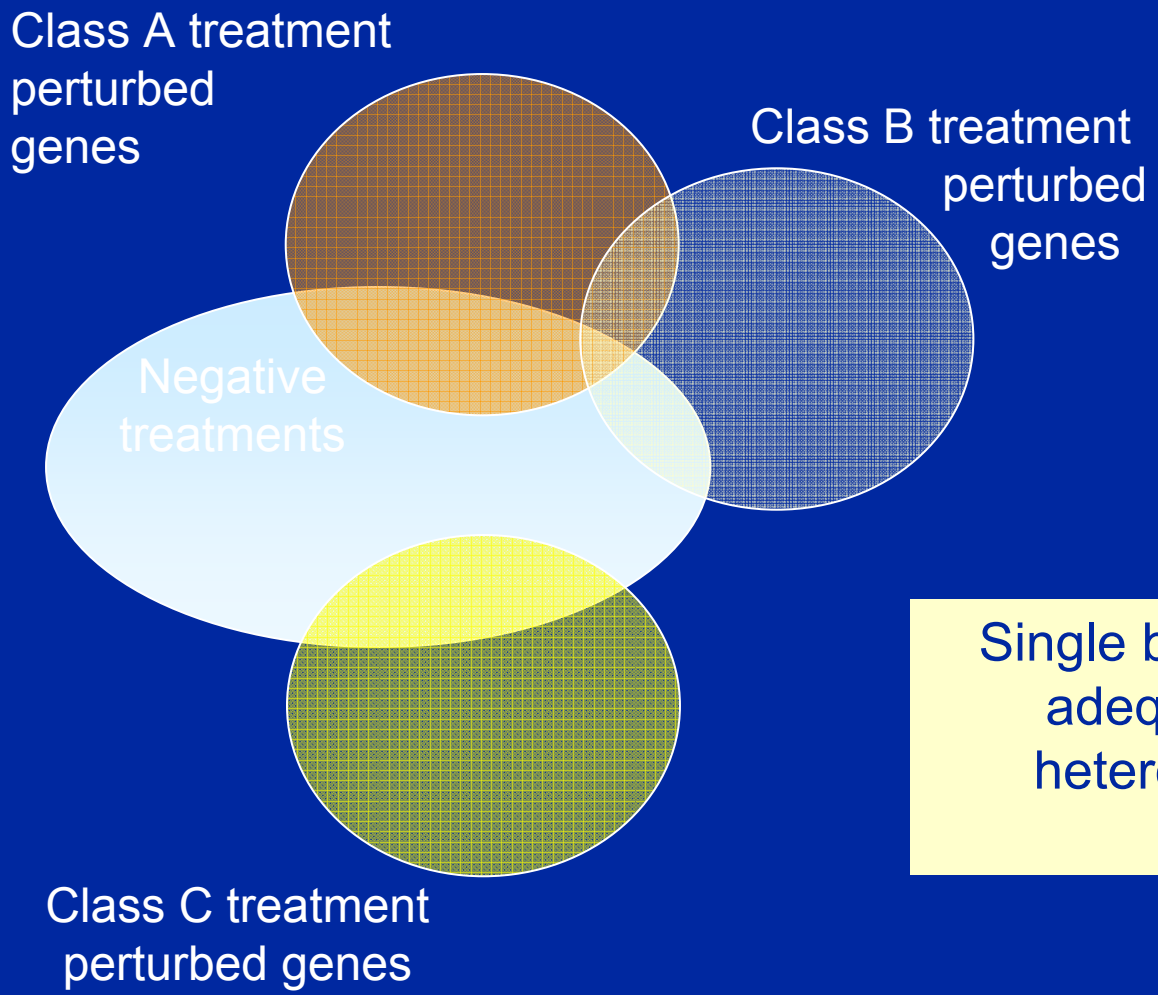


Data Provided by Iconix Biosciences

Traditional Biomarker Discovery Approach



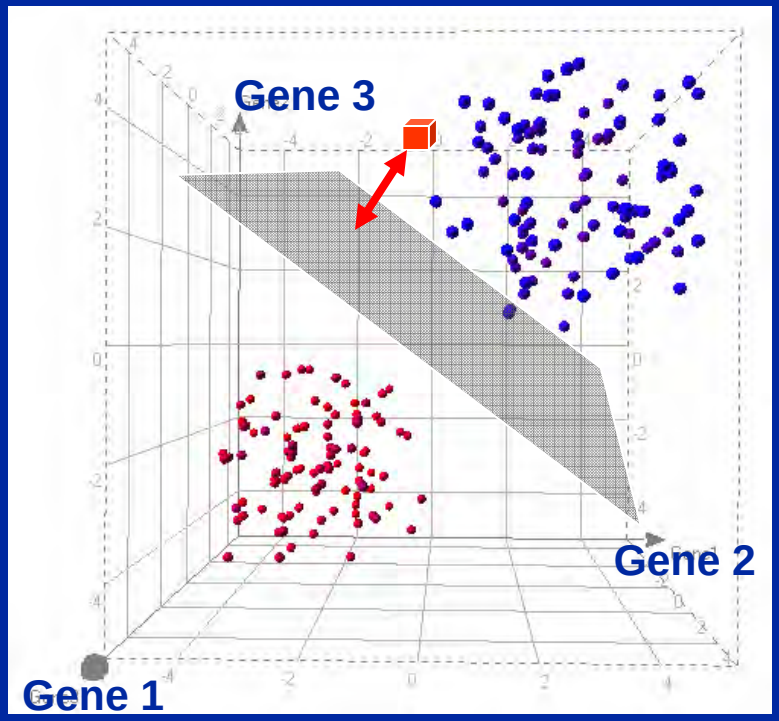
Contemporary Biomarker Discovery Approach



Single biomarker genes do not adequately represent the heterogeneous etiology of pathology

Linear Classification Algorithms

Algorithm attempts to find a linear separation between two classes in multi-dimensional gene space



Weights determine the orientation of the hyperplane,
Bias determines the position along the axis

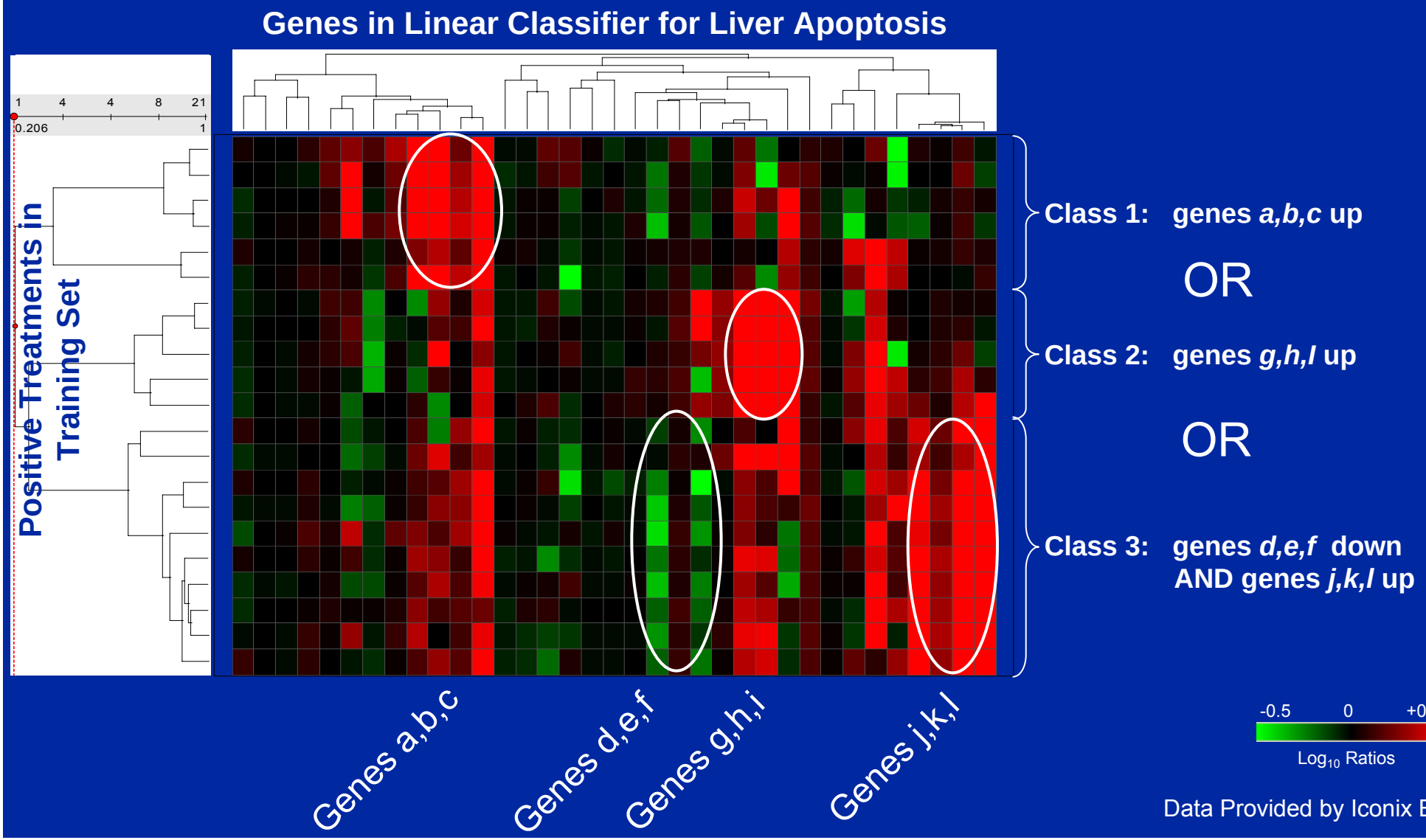
- **Log ratios for genes : x_1, x_2…x_n**
- **Associated weights : a_1, a_2…a_n**

$$S = \sum a_i x_i - b$$

S= Scalar Product and b = Bias

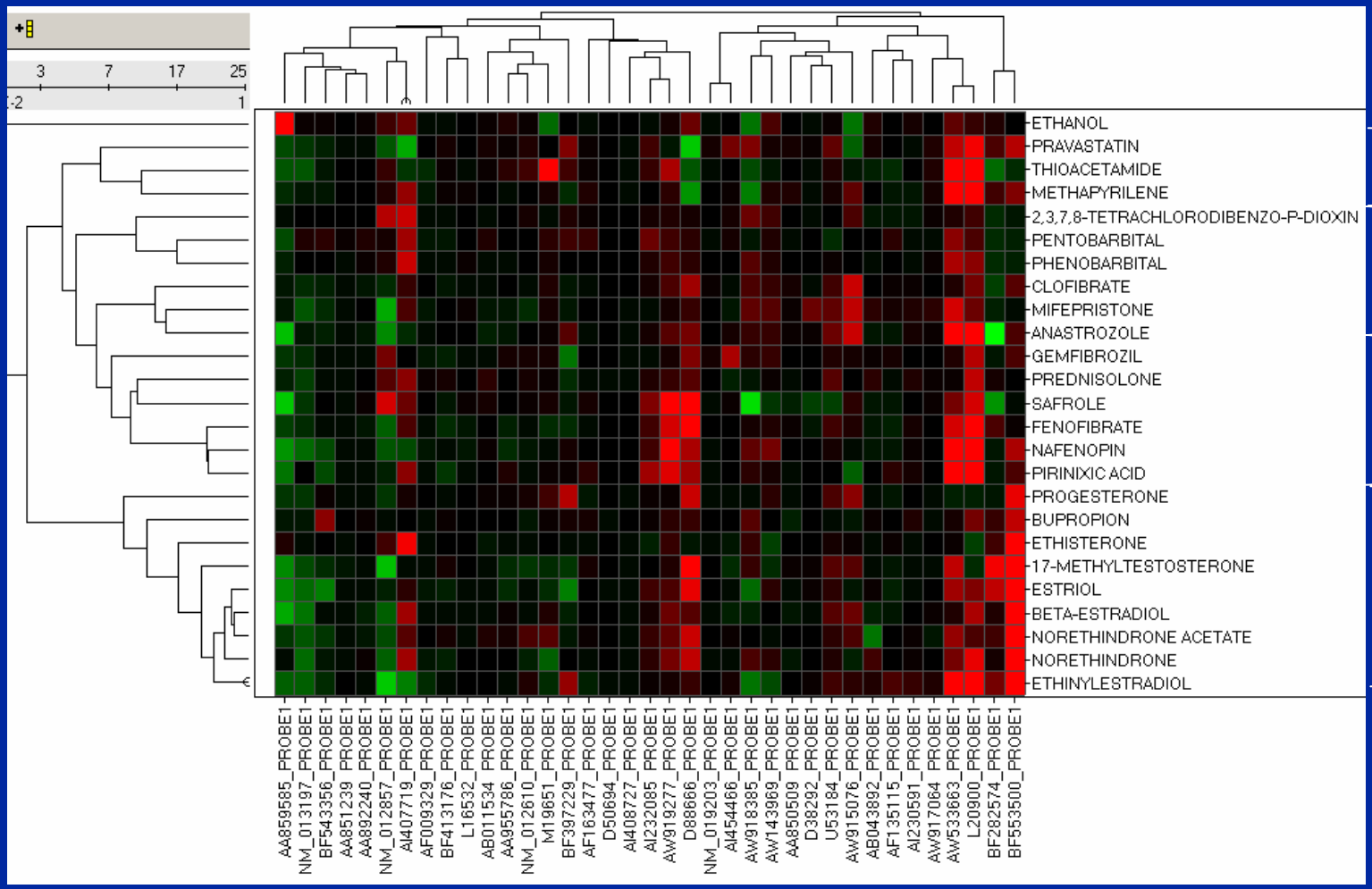
Interpretation:
If S>0 = True (in class)
If S<0 = False (not in class)

Linear Classifiers Use Multiple Genes to Account for Heterogeneous Classes



Data Provided by Iconix Biosciences

Chemicals with Similar Mechanism of Action Have Similar Profiles Based on Clustering of Genes



Toxicity and Regenerative Proliferation

Xenobiotic Receptor Mediated

PPAR Mediated

Hormonal Mediated

Genes in Linear Classifier for Non-Genotoxic Hepatocarcinogen

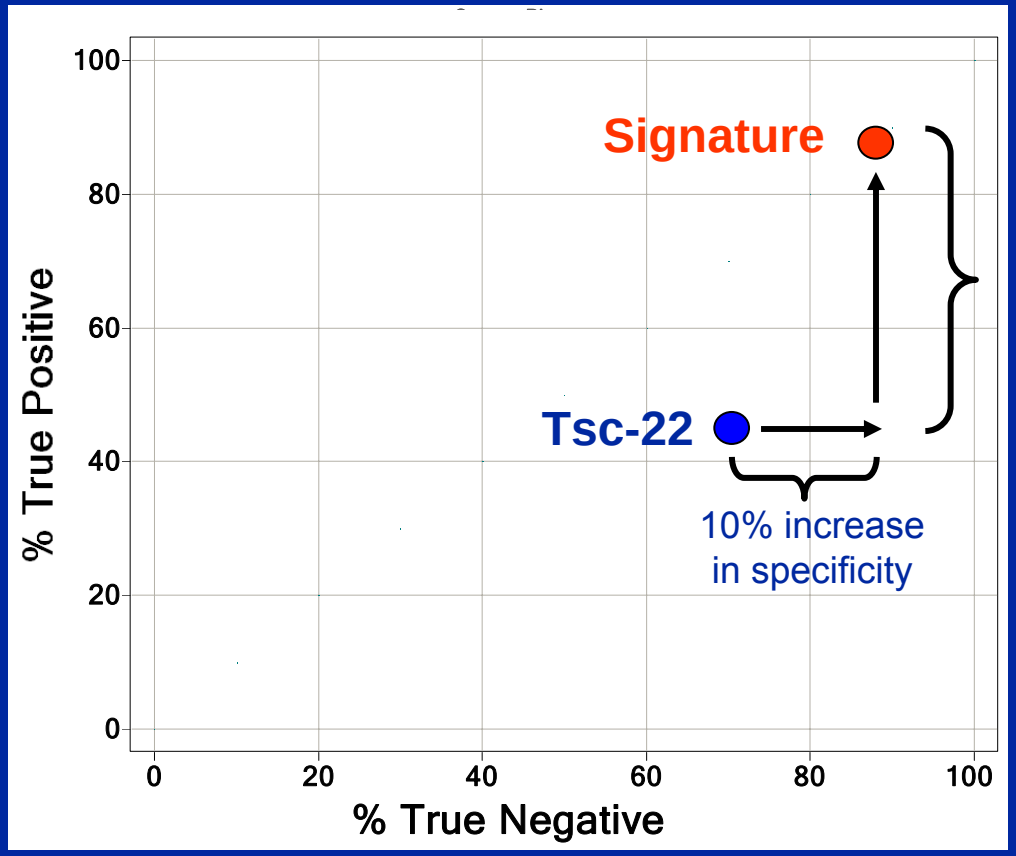
-0.5 0 +0.5

Impact (Log_{10} Ratios x weight)

Data Provided by Iconix Biosciences

Increased Accuracy of Multi-Gene Models vs Single Genes for Prediction

Signature to predict non-genotoxic hepatotumorigens



Based on independent test against 47 compounds

Anchoring Gene Expression Data to an Endpoint of Interest

- **Intrinsic Endpoints**
  - Expression data anchored to phenotype measured in the same sample

    In vivo: histopathology, clinical chemistry, organ weight, etc

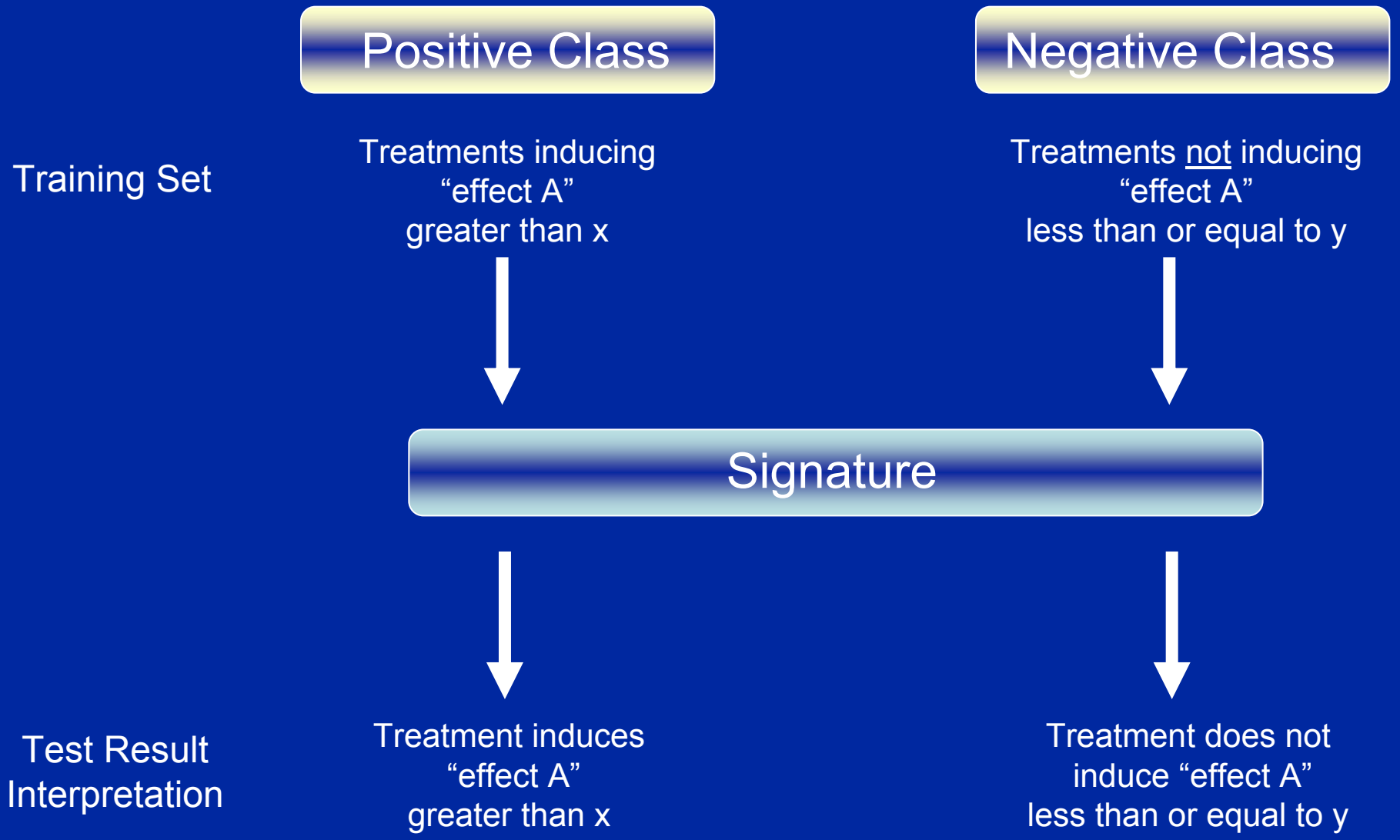
    In vitro: biochemical or structural change, cell size, shape, etc

- **Extrinsic Endpoints**
  - Expression data anchored to phenotype of treatment or compound determined elsewhere (i.e. literature)

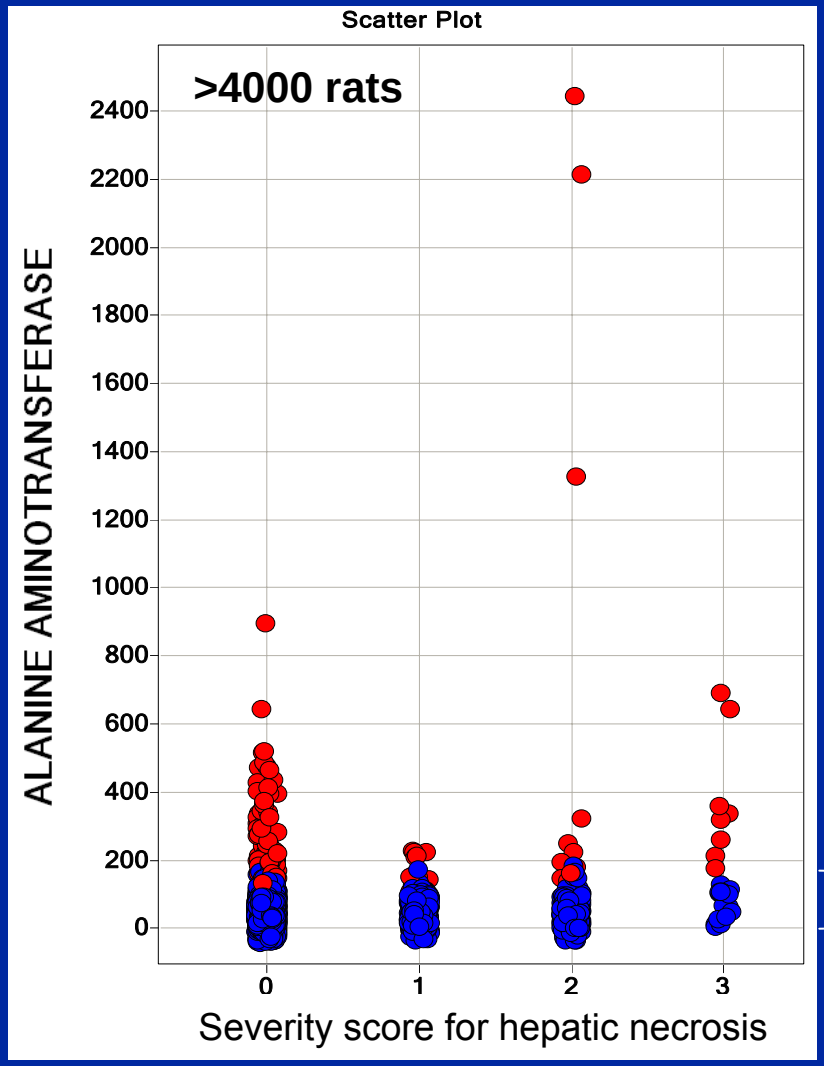
    In vivo: carcinogenicity, pharmacology

    In vitro: phospholipidosis, cholestasis, DNA damage

Classification Rules Determine Sensitivity and Specificity of Biomarker



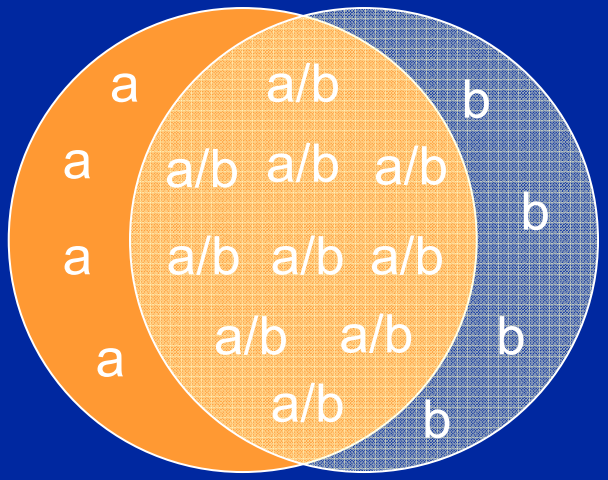
When Signatures Aren't What You Think They Are: Check your Classification Rule



Above normal range (control avg + 2SD)

Within normal range (control avg + 2SD)

When Signatures Aren't What You Think They Are: Check your Training Set for Confounding Variables



b = Pathology inducers
(Liver tumors)

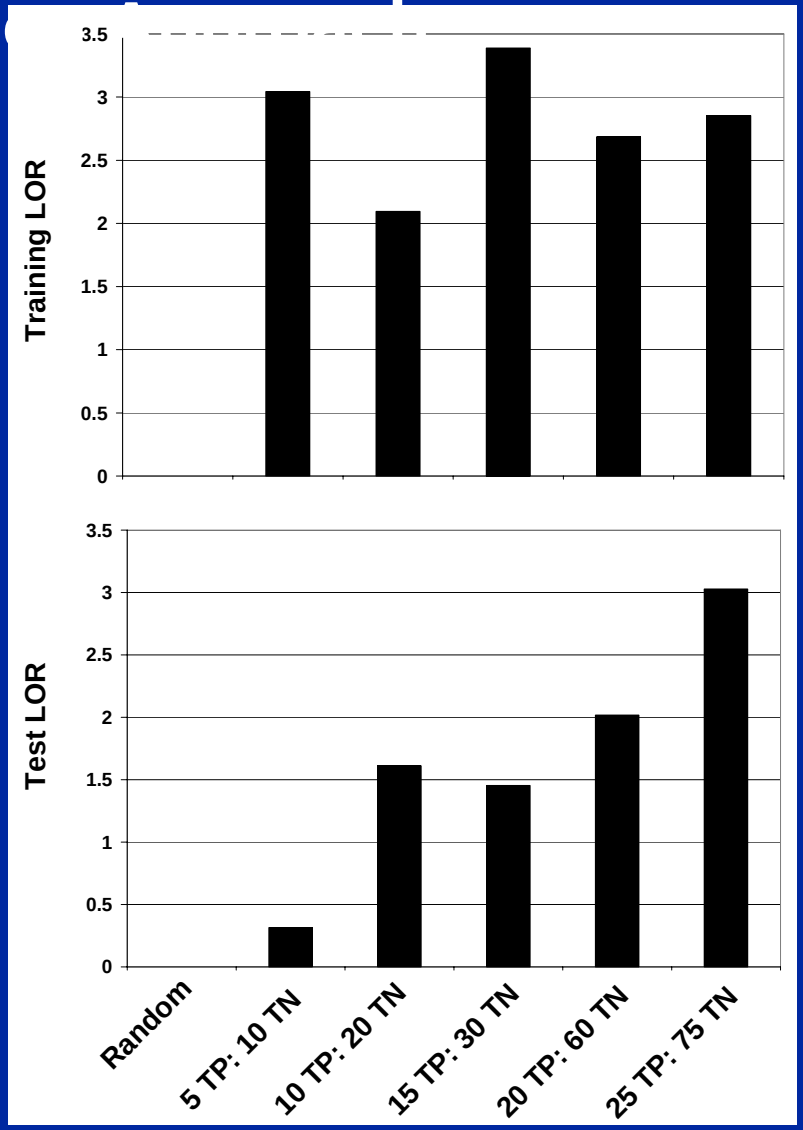
a = Pharmacological effect
(Nuclear receptor agonist)

Training set may represent distinct but correlated variables that dominate expression changes

When Signatures Aren't As Good As You Think They Are:
Check your Validati

Log Odds Ratio
Based on Cross Validation

Log Odds Ratio
Based on Independent Validation
(47 chemicals)



Estimated Accuracy

Real Accuracy

Final Thoughts

- Concentrate on validation, not discovery
- Don’t ignore confounding variables when interpreting data
- Predictor only as good as the training set from which it was derived (Size and diversity matters)
- Like other measured endpoints, predicted effects can be secondary in nature or not treatment related
- Classifier is only as accurate, but not more accurate, than the gold standard to which it is anchored
- Prediction is harder than originally thought
- Be realistic: “All models are wrong, some are useful” – *George Box*